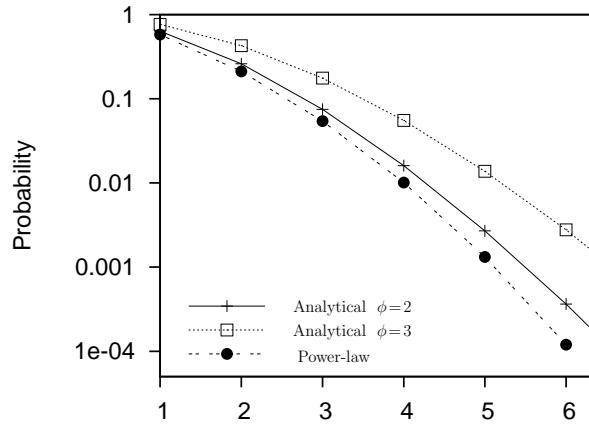
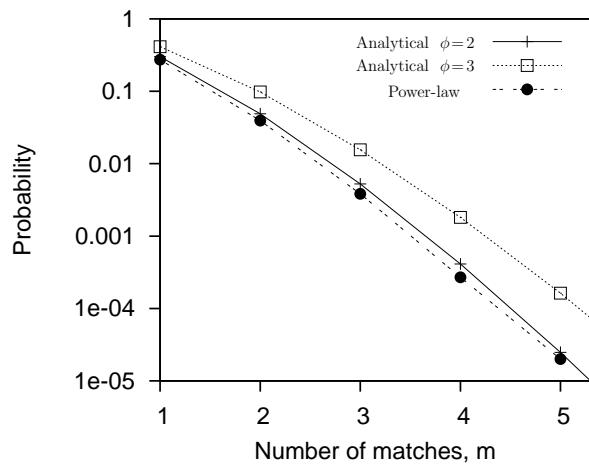


Gene cluster statistics with gene families:
Supplementary Figures

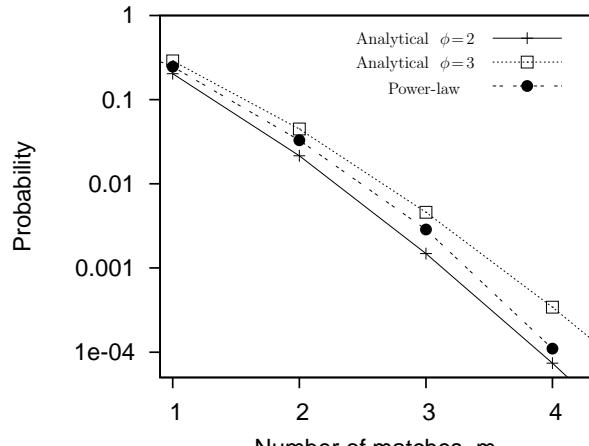
October 20, 2008



(a)

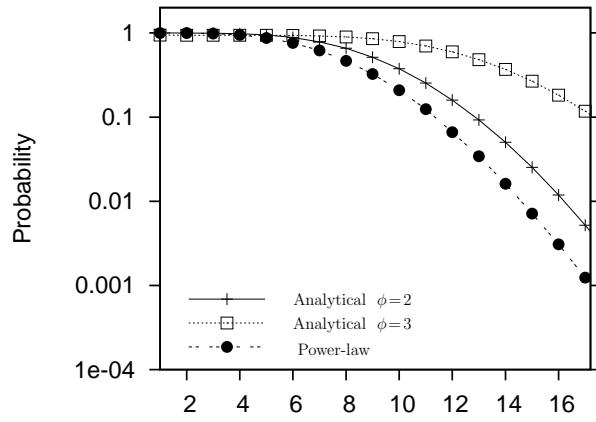


(b)

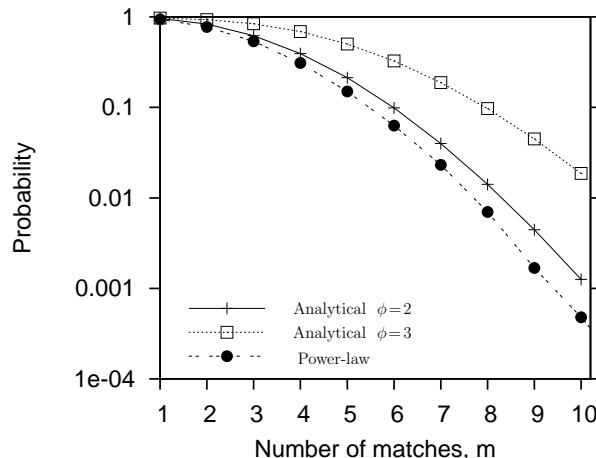


(c)

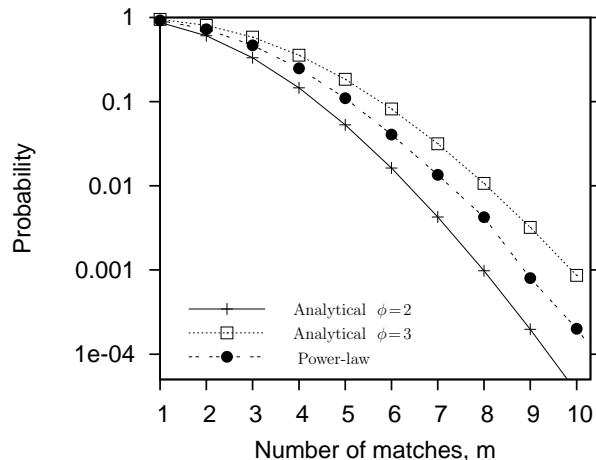
S 1: Comparison of orthologous cluster probabilities for power-law distributed and fixed-gene family sizes. The probabilities of observing at least one cluster of size m in a window of size $r = 50$ when genome size (a) $n = 5000$ (b) $n = 14000$ (c) $n = 22000$.



(a)

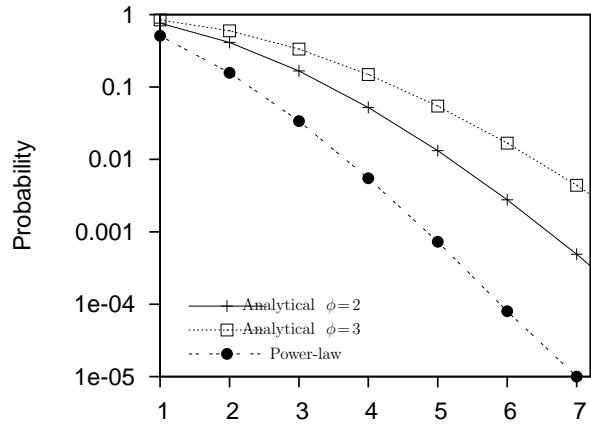


(b)

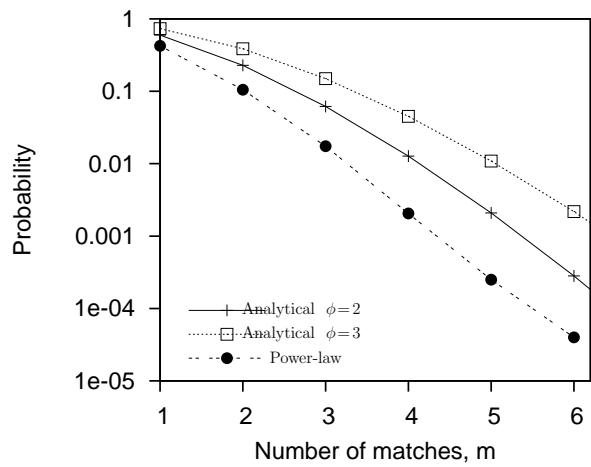


(c)

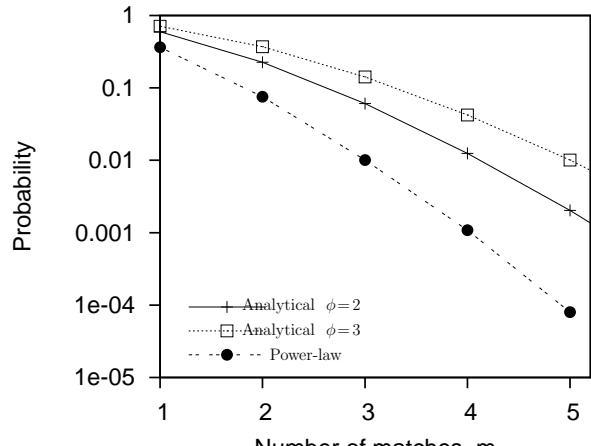
S 2: Comparison of orthologous cluster probabilities for power-law distributed and fixed-gene family sizes. The probabilities of observing at least one cluster of size m in a window of size $r = 150$ when genome size (a) $n = 5000$ (b) $n = 14000$ (c) $n = 22000$.



(a)

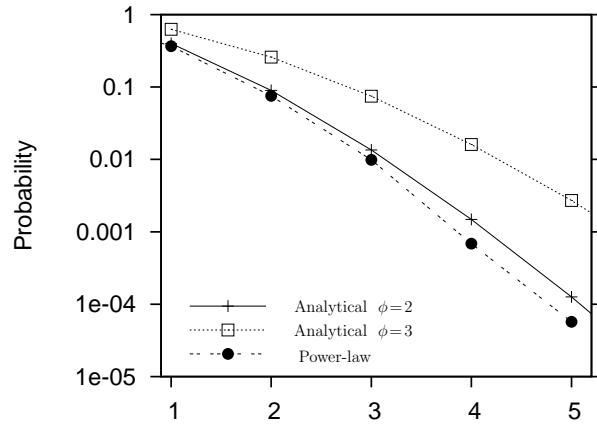


(b)

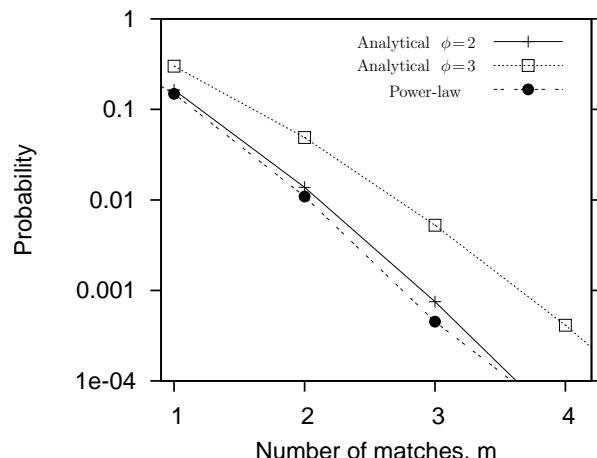


(c)

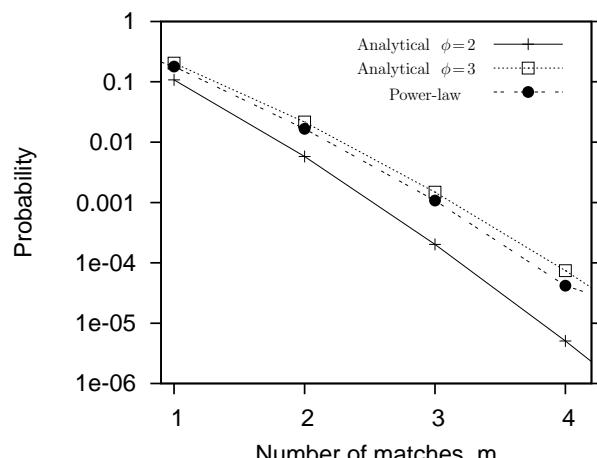
S 3: Comparison of orthologous cluster probabilities of two genomes of different sizes for power-law distributed and fixed-gene family sizes. The probabilities of observing at least one cluster of size m when (a) $n_1 = 5000, n_2 = 14000, r_1 = r_2 = 100$ (b) $n_1 = 14000, n_2 = 22000, r_1 = r_2 = 100$ (c) $n_1 = 5000, n_2 = 22000, r_1 = r_2 = 100$.



(a)

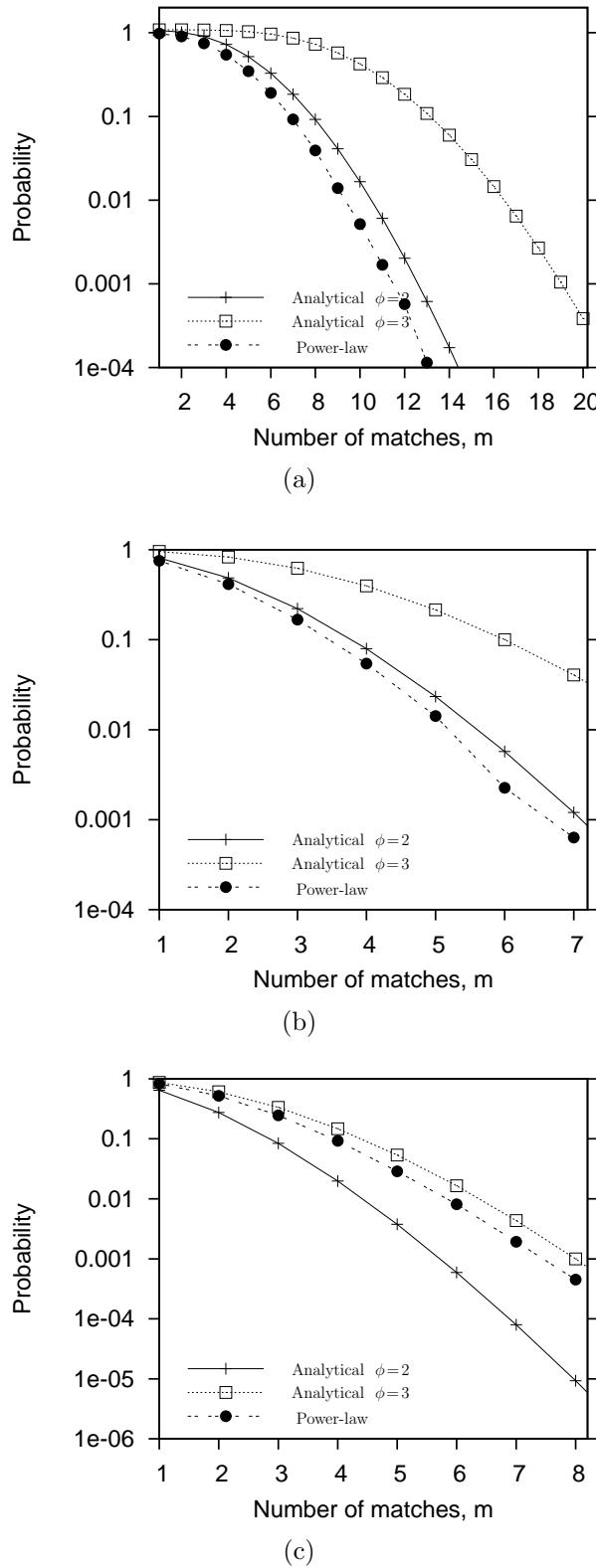


(b)

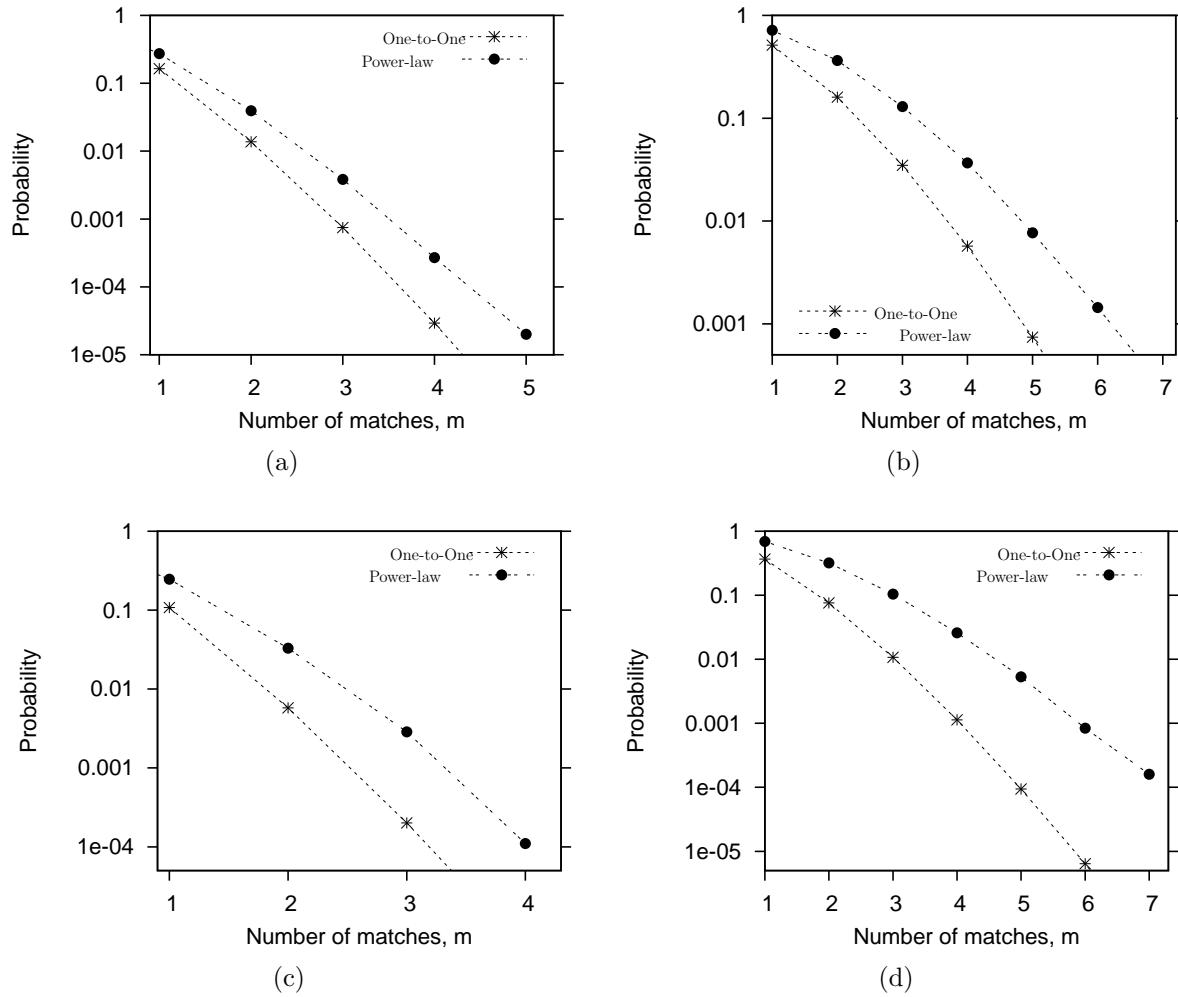


(c)

S 4: Comparison of paralogous cluster probabilities for power-law distributed and fixed-size gene families. The probabilities of observing at least one cluster of size m in a window of size $r = 50$ when genome size (a) $n = 5000$ (b) $n = 14000$ (c) $n = 22000$.



S 5: Comparison of paralogous cluster probabilities for power-law distributed and fixed-size gene families. The probabilities of observing at least one cluster of size m in a window of size $r = 150$ when genome size (a) $n = 5000$ (b) $n = 14000$ (c) $r = 150$.



S 6: Comparison of orthologous cluster probabilities for power-law distributed gene families and the one-to-one homology model. The probabilities of observing at least one cluster of size m in a window of size r given a genome of size n when (a) $n = 14000, r = 50$ (b) $n = 14000, r = 100$ (c) $n = 22000, r = 50$ (d) $n = 22000, r = 100$.